

Language Assessment



BERMUTU

Better Education through Reformed Management
and Universal Teacher Upgrading

Ministry of National Education
Directorate General of Quality Improvement of Teachers and Education Personnel
CENTER FOR DEVELOPMENT AND EMPOWERMENT OF
LANGUAGE TEACHERS AND EDUCATION PERSONNEL

2009

Language Assessment



BERMUTU

Better Education through Reformed Management
and Universal Teacher Upgrading

Ministry of National Education
Directorate General of Quality Improvement of Teachers and Education Personnel
CENTER FOR DEVELOPMENT AND EMPOWERMENT OF
LANGUAGE TEACHERS AND EDUCATION PERSONNEL
2009

Writers
Mansyur Bennu
Agus Purnomo

Reviewer
Fathur Rohim

PREFACE

Center for Development and Empowerment of Teachers and Education Personnel (CDELTEP) or Pusat Pengembangan dan Pemberdayaan Pendidik dan Tenaga Kependidikan (PPPPTK) Bahasa is in charge of promoting the quality of language teachers and school principal, school supervisor, and so forth. Hence, the Center takes part in the project of Better Education Through Reformed Management and Universal Teacher Upgrading (BERMUTU) in order to multiply their competencies and professionalism.

As a government institution that is professionally managed, PPPPTK Bahasa provides quality education services aligned with education reform and globalisation demand projected by Education for All (EFA). Likewise the institution develops Teacher Competency Standards inclusive teaching materials as a means of achieving the required competencies.

In the framework of the Minister National of Education Decree Number 14 year 2005 on Teacher and Lecturer, the Center, in an effort to generate competent and professional teachers, organizes various training activities to fulfill specific competency standards and certification programs. Therefore, the development of these learning materials are expected to be a useful resource for teachers.

Finally, constructive criticisms for further materials improvement are welcome and can be sent to PPPPTK Bahasa, Jalan Gardu, Srengseng Sawah, Jagakarsa, Jakarta 12640; Telephone (021) 7271034, Facsimile (021) 7271032, and email: admin@pppstkbahasa.net

Jakarta, September 2009
Center Director,

Muhammad Hatta, Ph.D.
NIP 19550720 198303 1 003

TABLE OF CONTENT

PREFACE	i
TABLE OF CONTENT	ii
CHAPTER I INTRODUCTION.....	1
A. Background.....	1
B. Objective.....	2
C. Indicators	2
CHAPTER II LANGUAGE ASSESSMENT	3
A. Terms Related to Language Assessment	3
B. The Characteristics of good Assessment	6
C. How to Assess Listening.....	9
D. How to Assess Speaking	19
E. How to Assess Reading.....	25
F. How to Assess Writing	30
G. Test Items Analysis.....	40
CHAPTER III CLOSING REMARK	47
REFERENCES.....	48
SUGGESTED READING	49

CHAPTER I

INTRODUCTION

A. Background

Language assessment is an integral part of teaching and learning process. It is actually one subject that most teachers are familiar with, yet it is also the one that often causes uncertainty and confusion among teachers. For one thing, many teachers still cannot grasp clearly and comprehensively on how to design a good assessment. Consequently, they just do what they think is 'correct' assessment instinctively without proper knowledge nor training. Hence, teachers are left to wonder and unsure if their ways of assessing or testing are appropriate, correct, or even valid as far as principles of assessment are concerned. This problem gets even more complicated when assessing is related with authentic assessment as it is relatively more subjective in nature compared to traditional assessment or test. For another, teachers still lack the skill needed to construct test items while the demand for teachers to be able to construct test items is growing and cannot be overlooked.

In light with the problem discussed above, this module of Language Assessment is designed. It seeks to provide teachers with sufficient, if not comprehensive, insight on assessment and practical guidance as well as ideas on how to conduct good language assessment in general, and testing in particular.

This module discusses some terms related to language assessment such as evaluation, assessment and authentic assessment, measurement, testing, and how to assess students' four language skills (Listening, Speaking, Reading, and Writing). Portfolio as part of authentic assessment, self assessment, learning log/journal, and test item analysis, are also discussed.

B. Objective

This module is intended to better equip teachers in the BERMUTU training so that they can use it as reference, or additional materials to enhance their understanding and mastery of the main module.

C. Indicators

In order to achieve the objective, a set of indicators of competence has been determined. They are:

- Describing the terms: testing, evaluation, measurement, assessment, and authentic assessment.

- Describing the criteria for good assessment.

- Constructing listening assessment.

- Constructing speaking assessment.

- Constructing reading assessment.

- Constructing writing assessment.

- Analyzing multiple-choice test items.

- Analyzing essay test items.

CHAPTER II LANGUAGE ASSESMENT

Hello fellow teachers, after having read the outline in chapter I, surely we have grasped general description about the module. Now, in Chapter II we are going to discuss: firstly, a number of important terms related to assessment; secondly, techniques of assessing listening, speaking, reading, and writing; thirdly, other types of assessment; and finally, test items analysis.

Please read carefully and do the exercises that follow.

A. Terms related with Language Assessment

Many people confuse the terms of assessment, measurement, and test. As a result, they use them interchangeably. Below are definitions and descriptions of the terms. The term 'authentic assessment' will also be described.

1. Assessment

Below are two definitions cited in Airasian (1991) and Nunan (1991).

- ❖ Airasian (1991:5-7) defines assessment as a general term which includes all the ways of information is gathered
- ❖ The set of processes through which we make judgments about learner's level of skill and knowledge" (Nunan, 1990).

2. Evaluation

Evaluation is an act of identifying to see whether or not a program which has been planned beforehand is accomplished, worthy, and effective. Evaluation involves value judgement. Evaluation is actually the application of a variety of assessment ways and tools in order to gather information as to how far the learning objectives or students' competence have been achieved. **(Curriculum Center, Classrom Assessment Guide, 2003).**

Similarly, Airasian (1991) points out that evaluation involves making judgments (concerning what is desirable) based on that gathered material.

3. Test

Experts define 'test' differently as follows.

- ❖ The smaller component of the decision-making process-is a measurement tool designed to elicit a specific sample of behavior (Bachman 1990:21; Linacre 1993:31).
- ❖ Language tests are an operation of the construct(s) that the tester is seeking to measure (Bachman 1990; Davies 1990:2)
- ❖ A means of checking learning that has taken place with respect to a specified teaching content or input, often by means of certain task. The results are usually concrete and can be expressed quantitatively". (Vale and Feunteun, 1995)
- ❖ Any procedure for measuring ability, knowledge or performance.
- ❖ A global view of achievement of the teaching and learning process over a period of time. E.g. analysis of the success or failure of teaching approach, course book, pupil response, motivation, etc. (Vale and Feunteun, 1995).
- ❖ Examine process and product of the course
- ❖ A sample of behavior under controlled or specified conditions and aimed toward providing a basis for forming judgments. (Milagros Ibe, 1981).
- ❖ A test is a measurement instrument designed to elicit a specific sample of an individual behavior (Bachman,1990).

From the above quotations, we can conclude that

A **test** is narrow in focus, designed to measure a set of skills or behaviours at one point in time.

- **Assessment** is broader in scope and involves gathering information over a period of time. (might include formal tests, classroom observations, student self-assessments, or from other data sources).

- **Evaluation** applies assessment data that have been scored and analysed to make judgements, or draw inferences about students and educational programmes.

4. Authentic Assessment

As the old paradigm of teaching and learning like focus on language, teacher-centered, isolated skills, emphasis on products and the likes has shifted to new paradigm such as focus on communication, learner-centered, integrated skills, and emphasis on products, so has assessment in general.

Traditional forms of assessment, which adheres to the paradigm of (roughly) 'tests that test', have failed to provide the information on what students can really do in their second or foreign language. So, authentic assessment, which basically measures learner's ability to use language holistically in real-life situations and usually done in a period of time (Tests that also teach), has been proposed to challenge the traditional assessment.

In short, authentic or alternative assessment is any way of gathering information about peoples' language ability that does not follow to the standardization of traditional tests or assessment. Garcia and Person stated that..

.. alternative assessment consists of those efforts that do not adhere to the traditional criteria of standardization, efficiency, cost-effectiveness, objectivity, and machine scorability. Its main goal is to gather evidence about how students are approaching, processing, and completing 'real life' tasks in a particular domain."

The advantage of having authentic assessment is that the data compiled on individual students provide a clear picture of each student's development through their work. The teacher then can determine growth, area of weakness, and area of strength. He or she can also find out about his students' background, interests and others aspects necessary to develop learning process. In the end alternative

assessment can give teachers knowledge upon which they can reflect on, discuss, and assist students in the learning process.

Task

How would you differentiate among these terms: assessment, evaluation, test.

B. Characteristics of good Assessment

There are five characteristics of good assessment in general": practicality (is the test practical?), reliability (is the test consistent in its measurement?), validity (does the test measure what it is supposed to measure?), authenticity (is the test unbiased?), and wash back (is the test beneficial for learning process) (www.yesdil.com, downloaded on September, 5 2009).

1. Practicality

Harrison (1983) addresses the following questions (among others) concerning the practicality of a test.

How long will the test take?

Is any equipment needed (language laboratory, OHP)?

How long will it take to get the marking done?

How many people will be marked?

How will test materials be reproduced in quantity, and at what cost?

In short, test should be as economical as possible in time and in costs (including hidden time of time spent).

2. Reliability

A reliable test is consistent and dependable. It means that the test measures consistently. In other words, a reliable test will produce the same result if replicated. For example, the score of one student is similar regardless the time or day of the testing-assuming that there is no change in their ability. Or, in a student's writing of some narrative

story, it is probable that the characteristic of the writing will be similar if the same test is done, let's say twice, in a different week.

3. Validity

Validity is the most important principle as it addresses the 'heart' of what assessment is all about: it seeks to answer the question of "Does it measure what it is intended to measure?" Regarding validity, Hughes (, 2005) stated that a test that provides consistent measures of precisely the abilities that we are interested in is said to be *valid*.

Similarly, Harrison (1983) defines validity as the extent to which the test measures what it is intended to measure. He further mentions two kinds of validity: content validity and face validity. Content validity concerns what goes into the test. The test, he says, should represent as far as possible the areas to be assessed. So, before writing a test the teacher should make specification of areas to be tested based on the purposes of the assessment. Face validity concerns what teachers and students think of the test. In Hughes's words (2005) a test is said to have face validity if it looks as if it measures what it is suppose to measure. Hence we understand that face validity is concerned with the judgment of the students based on the appearance of the test.

4. Authenticity

In an authentic test

- the language is as natural as possible,
- items are as contextualized as possible,
- topics and situations are interesting, enjoyable, and/or humorous,
- some thematic organization, such as through a story line or episode is provided,
- tasks represent real-world tasks.

": (www.yesdil.com, downloaded on September, 5 2009)

5. Washback

Washback or backwash can be defined as the effect of testing to teaching and learning, as indicated by Hughes (2003): The effect of testing on teaching and learning is known as backwash, and can be harmful or beneficial.

Washback includes the effects of an assessment on teaching and learning prior to the assessment itself, that is, on preparation for the assessment.

Informal performance assessment is by nature more likely to have built-in washback effects because the teacher is usually providing interactive feedback.

Formal tests can also have positive washback, but they provide no washback if the students receive a simple letter grade or a single overall numerical score.

Classroom tests should serve as learning devices through which washback is achieved.

Students' incorrect responses can become windows of insight into further work.

Their correct responses need to be praised, especially when they represent accomplishments in a student's inter-language.

Washback enhances a number of basic principles of language acquisition:

intrinsic motivation, autonomy, self-confidence, language ego, interlanguage, and strategic investment, among others.

One way to enhance washback is to comment generously and specifically on test performance.

Washback implies that students have ready access to the teacher to discuss the feedback and evaluation he has given.

Teachers can raise the washback potential by asking students to use test results as a guide to setting goals for their future effort.

TASK

Analyze the test instruction below in terms of their validity. Are they valid? Does it measure what is intended to measure? Is there any other skill involved?

1. Write the conversation you have with a friend about the holiday you plan to have together.
2. You spend a year abroad, while you are there, you are asked to talk to a group of young people about life in your country. Write down what you would say to them.
3. Discuss the advantages and disadvantages of being born into a wealthy family.

C. How to Assess Listening

According to Brown (2004), we can derive four commonly identified types of listening performance, each of which comprises a category within which to consider assessment tasks and procedures.

Intensive. Listening for perception of the components (phonemes, words, intonation, discourse markers, etc.) of a larger stretch of language.

Responsive. Listening to a relatively short stretch of language (a greeting, question, command, comprehension check, etc.) in order to make an equally short response.

Selective. Processing stretches of discourse such as short monologues for several minutes in order to „scan” for certain information. The purpose of such performance is not necessarily to look for global or general meanings, but to be able to comprehend designated information in a context of longer stretches of spoken language (such as classroom directions from

a teacher, TV or radio news items, or stories). Assessment tasks in selective listening could ask students, for example, to listen for names, numbers, a grammatical category, directions (in a map exercise), or certain facts and events.

Extensive. Listening to develop a top down, global understanding of spoken language. Extensive performance ranges from listening to lengthy lectures to listening to a conversation and deriving a comprehensive message or purpose. Listening for the gist, for the main idea, and making inferences are all part of extensive listening.

The rest of the chapter deals with some principles of assessing listening and possible techniques.

1. Principles of assessing listening

In this section, some factors influencing the quality of listening assessment are considered.

a. Talking –not reading

Avoid reading aloud long written texts. Written texts lack most of the redundant features which are so important in helping students to understand speech. Most of us have experienced how much more difficult it is to follow a lecture which is read aloud than one which is given from brief notes.

It is a good idea to use pictures to help you give a talk for listening comprehension. Describe a picture to the class or use a series of pictures to tell a story or to describe a process.

b. Reading texts aloud

Although it is easy to advise giving talks from notes (as oppose to reading he talk aloud), in practice it is difficult to do this in a foreign language. Some teachers may not feel confident enough to give a talk spontaneously. As a result it is often necessary to read aloud a written text. If you do decide to read a talk aloud, try to make the following appropriate changes to the written text first.

Rewrite most of the complex sentences, making sure that the sentences for reading aloud are fairly short.

Rewrite the talk, and repeat the important points. Try restating the key points in different language.

When you read the written text aloud, pause slightly longer than normally at the end of clauses and sentences. Longer pauses are far better than reading the text very slowly.

c. Using Recording

It is true that it is much harder to understand speakers if we can't see them. Thus, talks and conversations recorded in cassettes tape are much harder test of understanding than those given in real life.

However, cassette tape recording has the following advantages:

They help make a listening test more reliable. If the teacher were to give the same talk on different occasion, it would vary a lot.

It is possible to use the (recorded) voices of native speakers or fluent speakers of English.

It is possible to play recording of conversations involving two or more speakers-something which is far better than the teacher trying to read aloud the voices of different speakers.

2. Possible techniques

All techniques of testing listening can be classified into: listen and do; listen and draw; listen and choose; listen and arrange; and listen and write.

a. Listen and do

This technique is also called 'following instruction', which is self explanatory, where the students follow the teacher's instructions, such as: stand up! Come here!; Sit down!.

In order for the students to enjoy the test, the teacher can ask them to play a game called, 'Simon says'. The purpose of the game is that only when the teacher begins his instruction with 'Simon says' that they have to follow the instruction.

b. Listen and draw

First of all, tell the students that they do not have to be a good painter, and that the beauty of the picture they will produce does not count. After that, read them a text and ask them to draw a picture based on the text.

Ur (1984) uses the following story for elementary level students.

There is a table in the middle of the picture and a cat is under the table. He is a white cat. Near the table is a chair. There is a very fat boy sitting on it. He is very fat indeed, and very happy, because there is a big cake on the table, and he's going to eat the mouse which is under the fat boy's chair

c. Listen and choose

This technique requires students to choose one out of several pictures that best matches the description they hear from the teacher. Puzzle is one version of this technique. The teacher can choose a description of an animal such as: 'I has four legs, it is big. You can ride it. What is it?' If the pictures available (for them to choose) are a goat, a horse, a cat, and a rooster, the students are expected to choose the picture of a horse.

Blundell and Stokes (1984) exemplify this technique using a dialogue as follows.

Fiona : Hello.

Stuart:: Hello.

F : Oh. Is that Stuart?

S : Yes.

F : Hello Stuart. It's Fiona.

S : Oh. Hi Fiona.

F : Hi. Erm...is Judy there by any chance?

S : No. I'm sorry she's just popped out to the shops.

F : Oh dear. Erm ... could you possibly leave a message?

S : Yes. Yes. Just a second, let me get a piece ... bit of paper.

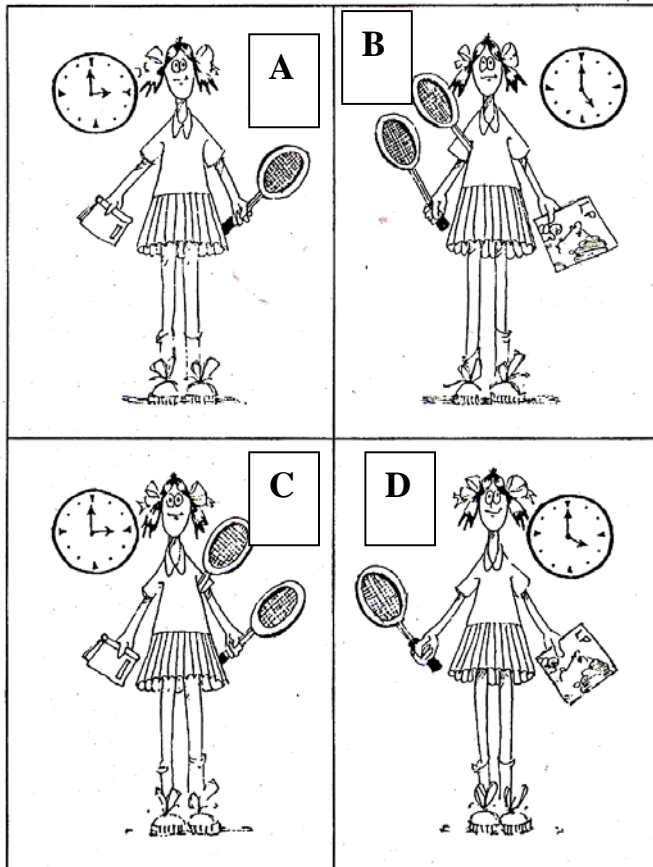
F : Thank you.

S : OK.

- F : Er.... The thing is we've arranged to play tennis this afternoon (Mm-mm) at 3 o'clock (Yes) ... erm ... but I've got a problem because the string on my racquet's broken (Mm-mm) but I think that Judi's got an extra racquet (Yes I think she has) and so I was wondering if you could ask her to bring the extra one along.
- S : Yes. OK. I'll do that.
- F : OK and .. er.... oh yes one other thing. She borrowed a book from me (Mm-mm) and I think she's probably forgotten all about it. I wonder if you could possibly remind her to bring that along as well.
- S : She knows what it is , does she?
- F : Yes, yes. It's a novel.
- S : Yes. OK. So bring extra racquet and,,, er...the book that she borrowed.
- F : That's right. (OK) 3 o'clock.
- S : I'll tell her.
- F : Thanks very much Stuart.
- S : OK. Cheerio.
- F : Bye.
- S : Bye.

Below is the student's worksheet.

Which picture best represents the message?
Put a tick beside the correct one.



Multiple choice, which is the most frequently used technique in testing listening, can actually be classified as 'listen and choose', because while listening to a statement, a dialogue, or a monolog, the students are required to choose one of the four options provided. Hughes (2005) suggested that if multiple choice is to be used, the alternatives (options) must be kept short and simple.

d. Listen and arrange

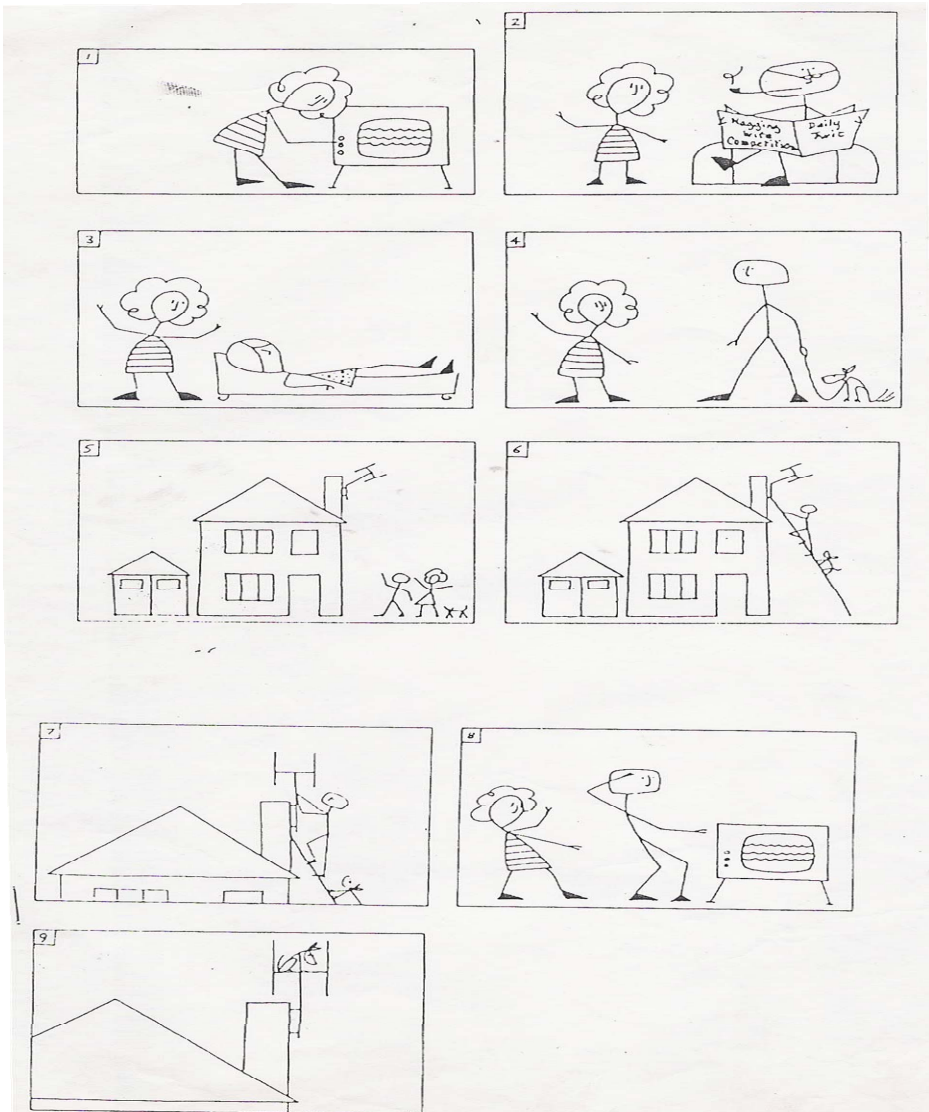
Like 'listen and choose', 'listen and arrange' needs a number of pictures. The difference is that the former requires students to choose only one picture, whereas the latter requires them to arrange all the pictures to represent the text they hear from the teacher or recording.

The more pictures to be arranged by the students, the more difficult the task will be.

Ur (1984) a monolog as the tape script' and nine pictures for students to arrange as follows.

Yesterday, mum wanted to watch television, but when she turned it on, she found it wasn't working. All she could get was some wavy lines. She asked her husband if he could do something about it, but he was busy reading his newspaper and smoking his pipe. She asked her daughter – but her daughter was having a rest and was too lazy to get up. Then she asked her son, Kevin. 'O.K., mum', he said. 'Let's go and have a look at the aerial first.' They went outside with the dog and looked up at the roof. 'Look mum,' said Kevin. 'All I need to do is straighten the aerial. Wait a moment.' He brought a ladder and climbed up on it, not noticing that his dog was climbing after him. Up on the roof he carefully straighten the aerial. Back down in the house they tried out the television again and found that was still showing only wavy lines. 'My goodness', said Mum. 'What is the matter now?' 'I don know,' said Kevin. 'By the way, where is the dog?' The dog, of course, was still sitting up on the aerial; that was what was causing the interference.

The student's task is to arrange these pictures while listening to the text above read by the teacher).



e. Listen and write

The term 'write' in this technique is not writing a paragraph, but rather a sentence, a phrase, or even a word as the answer to a particular question. A sample task of this technique is when the teacher reads a complete text, and at the same time, the students complete the same text, some words of which have been

removed/deleted. In other words, the students complete a gaped text while listening to the complete version of the text.

Another example given by Richard (1990) using a conversation as the 'tape script' as follows.

How long have these people been in Britain? Listen and write the number of weeks, months or years they have been here.

1. A. How long have you been here?
B. Let's see. I've been here for about six and a half weeks now.
2. A. Have you been living in Briton long?
B. Yes, I came here sixteen years ago.
3. A. Have you lived here long?
B. For about three and a half months now.
4. A. And you've been here for about three months?
B. No. I've been here for nine months now.
5. A. So when did you first came to the UK?
B. Oh, let me see now. Mm, it was about 26 years ago.
6. A. Did you arrive in UK recently?
B. Yes, I've only been here four and a half weeks.
7. A. How long have you been living in England?
B. Mm. This is my fifth year now.
8. A. Have you been living here for very long?
B. No, only five weeks.

And the students' worksheet is`as follows.

Activity

How long have these people been in? Britain Listen and write the number of weeks, months, or years they have been here.

- | | |
|----------|----------|
| 1. _____ | 5. _____ |
| 2. _____ | 6. _____ |
| 3. _____ | 7. _____ |
| 4. _____ | 8. _____ |

Another way of classifying the techniques of testing listening is as follows.

f. Simple instructions

Writing simple instruction for listening comprehension tests is usually a straight forward task. For example, we may put certain objects on a table and ask individual students to do thing with them.

g. Statements, questions, and short conversations

These require students to listen to statements, questions, and short conversations and then to choose the correct written option from a choice of four. Examples are given here of each type of them.

Statements

Students hear:

Tuti wouldn't be late so much if she knew she'd we punished.

Students read:

- A. Tuti is often late and is punished.
- B. Tuti is often late but is never punished.
- C. Tuti is rarely late and is never punished.
- D. Tuti is rarely late and thus isn't often punished.

Questions

Students hear:

How did you do it?

Students read:

- A. No I don't.
- B. I'm fine, thanks.
- C. About half an hour.
- D. By loosing the screw.

Short conversations

Students hear:

Man : Can I have an appointment with Dr Lawson tomorrow?

Woman : Let me see. He usually gets here at nine in the morning and he's free for ten minutes then. Otherwise it'll have to be ten-twenty.

Man : In that case, I'll be waiting for him as soon as he arrives.

Students read:

At what time does the man want to see Rd Lawson?

- A. 9.00 a.m
- B. 9.10 a.m
- C. 10.00 a.m
- D. 10.20 a.m

h. Following directions

First of all, students are asked to get ready with their maps. They then follow directions from the teacher in order to find the place referred to.

i. Radio reports

If you can listen to any radio programs in English you may be able to record the following types of reports for use as listening comprehension material.

Weather and traffic reports

Reports on today's events

News reports

Sports reports and results

j. Talks and lectures

If you give talks and short lectures for listening comprehension, remember that the test should not become a test of memory. Test students' understanding of the important points in the talk.

D. How to Assess Speaking

Weng (2006) points out five principles of assessing speaking skills as follows.

Ensure speaking contexts are as authentic as possible, esp. in communicative testing.

Put the candidates at ease prior to the real test

Remain as non-intrusive (for examiners) as possible esp. in group discussions.

Ensure scoring reliability.

Create speaking contexts that give maximum opportunities for candidates to speak.

Generally, according to Hughes (2005), there are three techniques of testing speaking: interview, interaction with fellow candidates (students), responses to audio- or video-recorded stimuli. In addition to examples suggested by Hughes, we also include one (problem solving) suggested by Heaton (1988).

1. Interview

The most common technique for the testing of speaking is the interview. However, there is at least one drawback of this technique. The relationship between the tester and the candidate is usually such that the candidate speaks as to a superior and is unwilling to take the initiative. As a result, only one style of speech is elicited, and many functions are not represented in the candidate's performance. Hughes (2005) suggests a way to overcome this problem by introducing a variety of elicitation techniques into the interview situation.

Requests for elaboration: What exactly do you mean?, can you explain that in a more detail?

Appearing not to understand: I'm sorry, but I don't quite follow you.

Invitation to ask questions: Is there anything you'd like to ask me?

Role play: Candidates can be asked to assume a role in a particular situation, such as: A friend invites you to a party on an evening when you want to stay at home and watch the last episode of a television serial. Thank the friend (played by the tester) and refuse politely.

Another example is: you want your mother (played by the tester) to increase your pocket money. She is resistant to the idea. Try to make her change her mind.

2. Interaction with fellow candidates

An advantage of having candidates interacting with each other is that they may feel more confident than when facing a dominant interviewer. As a result, they may perform better.

However, there is a problem, because the performance of one candidate is likely to be affected by that of the other(s). An insensitive candidate may dominate and not allow another candidate to show what he or she can do. To overcome this problem, Hughes (2005) suggests that the pairs should be carefully matched whenever possible. This technique may also involve more than two candidates.

3. Discussion

An example of this technique is to set a task that requires discussion between the two candidates. Tasks may require the candidates to go beyond discussion, and, for example, make a decision.

4. Role play

Role play can be carried out by two candidates with the tester as an observer. For some roles this may be more natural than if the tester is involved. It may, for example, be difficult to imagine the tester as 'a friend'.

5. Responses to audio audio-or video-recordings

According to Hughes (2005), universality of elicitation procedures can be achieved by presenting all candidates with the same computer generated or audio-/video-recorded (to which the candidates themselves respond into a microphone).

This technique can be economical where language laboratory is available, since large numbers of candidates can be tested at the same time. One obvious disadvantage of this technique is its inflexibility: there is no way of following up candidates' responses.

Below are examples taken from the ARELS (Association of Recognized English Language Schools).

Described situations

For example: You are walking through town one day and you meet two friends who you were sure had gone to live in the USA. What do you say?

Remarks in isolation to respond to

For example: The candidate hears, 'I'm afraid I haven't managed to fix that cassette player of yours yet. Sorry.'

6. Problem-Solving

One of techniques of authentic assessment that focuses on speaking is Problem Solving. In this technique, teacher creates imaginary or real problems designed in such a way that stimulates agreement and disagreement among students when they try to find solution of the problem(s). One example is as the following:

“Radioactivity from a nuclear power station accident will reach your area in a few hours. There is a small but very safe nuclear fallout shelter nearby, but there is room for only six people out of a total twelve. Which six people from the following list do you think would be most useful to save in the interest of future generation? List them in order of priority (Note: M=Male; F= Female).

- a marine biologist, aged 56 (F)
- a physicist, aged 25 (M)
- a famous musician, aged 38 (F)
- a farmer, aged 32 (M)
- an electrician, aged 49 (M)
- a mathematics teacher, aged 34 (F)
- a well-known footballer, aged 22 (M)
- a doctor, aged 63 (F)
- a university student of Sociology, aged 19 (F)
- a fireman, aged 33 (M)
- a factory worker, aged 28 (F)
- a garage mechanic, aged 27 (M)

Now, try to answer some questions below. Feel free to express your own ideas.

TASK

1. In your opinion, what language function is tested (or practiced) in this technique (for example: extending sympathy)?
2. Find other problem solving situations which relates closely to students life.
3. What are the benefits of having this kind of assessment technique?
4. How can teachers give score to students' speaking performance while having this technique?

Rating Scheme for Speaking

Accent

1. Pronunciation frequently unintelligible
2. Frequent gross errors and a very heavy accent make understanding difficult, require frequent repetition.
3. "Foreign accent" requires concentrated listening, and mispronunciation lead to occasional misunderstanding and apparent errors in grammar or vocabulary.
4. Marked "foreign accent" and occasional mispronunciations which do not interfere with understanding.
5. No conspicuous mispronunciations, but would not be taken for a native speaker.
6. Native pronunciation, with no trace of "foreign accent."

Grammar

1. Grammar almost entirely inaccurate except in stock phrases.
2. Constant errors showing control of very few major patterns and frequently preventing communication.
3. Frequent errors showing some major patterns uncontrolled and causing occasional irritation and misunderstanding.

4. Occasional errors showing imperfect control of some patterns but no weakness that causes misunderstanding.
5. Few errors, with no patterns of failure.
6. No more than two errors during the interview.

Vocabulary

1. Vocabulary inadequate for even the simplest conversation.
2. Vocabulary limited to basic personal and survival areas (time, food, transportation, family, etc.)
3. Choice of words sometimes inaccurate, limitations of vocabulary prevent discussion of some common professional and social topics.
4. Professional vocabulary adequate to discuss special interest; general vocabulary permits discussion of any non-technical subject with some circumlocutions.
5. Professional vocabulary broad and precise; general vocabulary adequate to cope with complex practical problems and varied social situations.
6. Vocabulary apparently as accurate and extensive as that of an educated native speakers.

Fluency

1. Speech is so halting and fragmentary that conversation is virtually impossible.
2. Speech is very slow and uneven except for short or routine sentences.
3. Speech is frequently hesitant and jerky; sentences may be left uncompleted.
4. Speech is occasionally hesitant; with some unevenness caused by rephrasing and grouping of words.
5. Speech is effortless and smooth, but perceptively non-native in speed and evenness.
6. Speech on all professional and general topics as effortless and smooth as a native speaker's.

Comprehension

1. Understands too little for the simplest type of conversation.

2. Understands only slowly, very simple speech on common social and touristic topics; requires constant repetition and rephrasing.
3. Understands careful, somewhat simplified speech when engaged in a dialogue, but may require considerable repetition and rephrasing.
4. Understands quite normal well educated speech when engaged in a dialogue, but requires occasional repetition or rephrasing.
5. Understands everything in normal educated conversation except for very colloquial or low-frequency items, or exceptionally rapid or slurred speech.
6. Understands everything in both formal and colloquial speech to be expected of an educated native speaker.

WEIGHTING TABLE

	1	2	3	4	5	6	(A)
Accent	0	1	2	2	3	4	
Grammar	6	12	18	24	30	36	
Vocabulary	4	8	12	16	20	24	
Fluency	2	4	6	8	10	12	
Comprehension	4	8	12	15	19	23	
Total							

Note the relative weighting for the various components.

E. How to Assess Reading

In this section, we shall discuss some criteria for selecting texts and possible techniques that can be used to assess students' ability to understand reading texts.

1. Selecting texts

Successful choice of text very much depends on experience of constructing tests, including text selection. It is nevertheless possible

to offer possible advice. For this purpose, Hughes (2005) suggests the following.

Keep specifications constantly in mind and try to select as representative a sample as possible. Do not select texts simply because they are readily available.

Choose texts of appropriate length. Expeditious reading texts may call for passages of up to 2,000 words. Detailed reading can be tested using passages of just a few sentences.

In order to obtain both content validity and acceptable reliability, include as many passages as possible in a test, thereby giving candidates a good number of fresh starts.

In order to test search reading, use passages which contain plenty of discrete pieces of information.

Avoid texts made up of information that may be part of candidates' general knowledge.

Assuming that it is only reading ability that is being tested, do not choose texts that are too culturally laden.

Do not use texts that students have already read.

2. Possible techniques

It is important that the techniques used should interfere as little as possible with the reading itself, and that they should not add a significantly difficult task on top of reading.

a. Multiple choice

The candidate provides evidence of successful reading by making a mark against one out of a number of alternatives.

Multiple choice take many forms, but their basic structure is that there is a stem and a number of options, one of which is correct, the other being distracters.

One problem with multiple choice is that, the chance of guessing the correct answer in a four-option multiple choice item is one in four, or twenty five per cent. Another problem with multiple choice is that, good ones are very difficult to write. Common faults are: more than one correct answer; no correct answer; there are clues in the options

as to which is correct (for example, the correct option may be different in length to the others). It is also very easy or candidates to communicate non-verbally.

Rewriting part of the paragraph

Sometimes it may be necessary to rewrite part of the passage or add to it in order to include information for another distracter. For example, in the following multiple choice- item , option C is clearly wrong as it does not relate to anything at all in the reading passage.

As I gazed at the peaceful scene, I suddenly noticed a yacht gliding slowly across the bay. It looked almost like a big ship.

What did the writer see as he looked in front of him?

- A. Big ship
- B. A yacht
- C. A helicopter

In such a situation, it is necessary for us to go back to this part of reading passage and add information which will allow us to write a plausible distracter.

As I gazed at the peaceful scene, I suddenly noticed a yacht gliding slowly across the bay. It looked almost like a big ship. Overhead, the sky was clear. There was no sign of the helicopter which had circled above me the day before.

What did the writer see as he looked in front of him?

- A. Big ship
- B. A yacht
- C. A helicopter

Avoiding matching words

Try to avoid items which only require students to match words in the text. In the next example students can easily choose the correct answer even if they cannot really understand what is happening in the story.

Henry was looking for his cat when he suddenly saw a large black dog several hundred meters in from of him.

When Henry was in the park, he suddenly saw.....

- A. His cat
- B. Nothing at al
- C. A large black dog

Writing grammatically correct options

It is important to make all options grammatically correct if you use incomplete sentences in a multiple-choice item.

Her favorite sports were

- A. Swimming and playing hockey
- B. Cycling and tennis
- C. Volleyball (grammatically unacceptable)

Writing options of equal length

Make all your options equal in length in the same multiple-choice item. Some teachers tend to make the correct answer longer than the others –usually because they want to make sure that it is absolutely correct.

b. True/false items

True / false items offer a very reliable way of testing a student's reading comprehension, provided that there are enough such items in a reading test. You will find it much easier and quicker, moreover, to write true/false items than multiple choice items, a large number of true/false items can be written on a fairly short reading text.

c. Short answer questions

These are questions which require the students to write down specific answers in spaces provided on the question paper. The technique is useful for testing both reading and listening comprehension.

d. Cloze

In the cloze procedure words are deleted from a text after allowing a few sentences introduction. The deletion rate is mechanically set, usually between every fifth and eleventh word (Weir, 1990).

Some teachers do not distinguish between blank-filling and cloze tests, using the term cloze for all blank-filling tests. Although cloze test are similar in several ways to blank-filling tests, there are basic difference between the two.

In an ordinary blank-filling tests, we decide which words we will delete from a text. In this way we can test the students' ability to understand specific meanings in the text. In a cloze test, however, we never choose which words we want to omit: we delete the words systematically.

It is important to let the students see the first sentence or two without blanks. This will give them an opportunity to get used to the topic and style of the passage.

e. Selective deletion gap filling

Due to the negative findings on mechanical deletion cloze, increasing support has developed for the view that the test constructor should use a 'rational cloze', borrowing Weir's term (1990). That is to select items for deletion based on upon what is known about language, about difficulty in text and about the way language works in a particular text. Linguistic reasoning is used to decide on decisions and so it is easier to state what each test is intended to measure.

f. Information transfer

In this technique, the information transmitted verbally is transferred to a non-verbal form, e.g. by labeling a diagram, completing a chart, or numbering a sequence of events.

This technique can` also be used in testing listening comprehension.

Note on scoring

In a reading assessment (or a listening assessment), errors of grammar, spelling or pronunciation should not be penalized, provided that it is clear that the student has successfully performed the reading task which the item tests.

Task

In pairs, choose one technique of assessing reading, and write 5 (five) items.

F. How to Assess Writing

According to Weir (1990), two different approaches for assessing writing ability can be adopted: indirect and indirect. In indirect approach, writing is divided into discrete levels, e.g., grammar, vocabulary, spelling and punctuation, and these elements can be tested separately by the use of objective tests.

To be able to operate the four language skills (listening, speaking, reading and writing) in the various function areas, it is necessary to be able to manipulate items from three levels of language. That is, to communicate, it is necessary to have an adequate vocabulary, to know basic items of English grammar and to be able to handle English sounds, stress and intonation (Eldowney, in Weir, 1990). This quotation clearly supports the indirect approach of assessing writing. In our view, Eldowney is right, but we may not rely on the indirect approach alone, because the ultimate purpose of writing is to produce a complete text. We then suggest that we apply both approaches, starting with the indirect.

Direct approach focuses on more direct extended writing tasks of various types.

Supporting Hughes idea, Johnson (2007) suggests such formats as persuasive letters, brochures, short essays, as examples of what he calls extended written response (a kind of authentic assessment). Other kinds of authentic assessment, according to him, are portfolio, project, and performance.

Below are possible techniques of assessing writing, involving samples tasks of both approaches. There is no clear cut between controlled and free writing. The terms are used only for text organization purpose.

1. Grammar and structure

Wherever possible, set the items in context. The disadvantage of using a passage for a progress or an achievement test is simply that you may be prevented from choosing certain areas of grammar on which to concentrate, especially if the passage is fairly short. When you choose a paragraph, you can test only the grammatical points which occur in that paragraph.

If you want to concentrate on a certain area of grammar, put the item into a short two-line dialogue. This is better than providing no context at all. Thus, the item

_____ a pen and a piece of paper.

- A. I like
- B. I'll like
- C. I'd like
- D. I'm liking

Becomes more meaningful when expressed as a reply to a request:

X : Can I get you anything?

Y : _____ a pen and a piece of paper, please.

- A. I like
- B. I'll like
- C. I'd like
- D. I'm liking
- E. Editing task

Editing task

The student is given a text containing a number of errors of grammar, spelling and punctuation and is asked to rewrite making all the necessary corrections. The advantage of this technique is that it equals part of the writing process. Also, this task may have a good wash back effect in that students may be taught and encouraged to edit their written work more carefully.

Error-recognition items

Use the errors your students make in their composition to write this type of items. In each item, four words or phrases are underlined and marked A, B, C, and D. Students must choose the underlined word or phrase which is incorrect.

(1) It was a terrible accident at an air-show held in West Germany yesterday.

A B C D

(2) There was wreckage everywhere and hundred of people were killed or badly injured.

A B C
D

Re-arrangement

Students are asked to unscramble sentences. They must write out each sentence, putting the words and phrases in their correct order.

(1) I ask her if
anyone/ she/ to help/ find/ us/ could/

(2) I made Ardi
His parents/ to/ the broken bicycle/ show/

By requiring students to re-arrange sentences, you encourage them to pay full attention to such grammatical markers as connectives (since, but, although, however, etc.) and pronouns.

- A. However, we decided to set out and try to climb the mountain.
- B. This time we had better climbing equipment and were determined to succeed.
- C. Several people in the group had tried this before but no one had managed to get to the top.
- D. There was a strong wind and it was raining very heavily.

2. Controlled writing

Control over what students write in an exam may range from very strict control over the grammar and forms of language used to less strict control over the functions and subject matter. The following types of tasks show the range of control and some of the forms it can take.

Sentence and paragraph completion

Some items require students to complete sentences or paragraph. In this way, writing is integrated with reading comprehension and becomes a more realistic task.

Most of the students in my class were rather lazy and did not enjoy the course. Some even stayed away quite often. Susi, however,

Other blank-filling items like the form of incomplete dialogues.

A: _____
B: Certainly, please do.
A: _____
B: Not at all. Just ask the operator to tell you how much it is.

Form-filling

This is a very relevant task for students; in modern life it is often necessary for students to fill in forms. For examples, students may be given an application form for a junior high school. They are asked to give details about themselves.

Choosing subjects

Choose subjects which within the experience of your students Avoid very general, abstract topics such as 'The countryside at night', 'the importance of time', etc.

Realistic writing task

Think about the likely context in which students will write. Most people write letters in their daily lives. Therefore, test your students' ability to write letters. More importantly, because In the school based curriculum, students are offered texts such as descriptive, narrative, recount, report, procedure and short functional texts, teachers are advised to consider these types of texts when setting task for their students.

Writing for a purpose

Always try to give students a purpose of writing. This is important for motivating students and encouraging them to produce good written work.

Writing for an audience

It is useful if students know for whom they are writing. Knowing who our readers are influences the way we write.

Writing Criteria (Weir, 1990)

ACCURACY

- Level 1 Grammar, vocabulary, spelling and punctuation maybe uncertain but what candidates write is intelligible and unambiguous.
- Level 2 Generally good control of grammar, vocabulary spelling and punctuation though some errors which do not destroy communication are acceptable.
- Level 3 Good control of grammar, vocabulary, spelling and punctuation. Any errors must not interface with communication
- Level 4 Standards of grammar, vocabulary, spelling and punctuation are consistently of a very high level.

APPROPRIACY

- Level 1 Use of language is broadly appropriate to the task, but no subtlety is expected. The intention of the writer can be perceived without excessive effort. Layout is generally appropriate but may show marked inconsistencies.
- Level 2 Use of language is in most respects appropriate to the task, and some adaptation of style to the particular context is demonstrated. The overall intention of the writer is clear. Layout, including handwriting, is generally appropriate.
- Level 3 Use of language is in most all respects appropriate to the task. There is clear evidence of the ability to adapt style to the particular context. The intention of the writer, both overall and in detail, is generally clear. Layout, including handwriting is generally appropriate.
- Level 4 Use of language is consistently appropriate to the task, context and intention. Layout is consistent and appropriate. Handwriting does not interfere with communication

RANGE

- Level 1 Severely limited range of expression is acceptable. Candidates may have labored to fit what they want to say to what they are able to say.
- Level 2 A fair range of language is used. Candidates are able to express themselves without gross distortion.
- Level 3 An extensive range of language is used. Candidates are able to express themselves clearly and without significant distortion.
- Level 4 Few limitations on the range of language available to candidates are apparent. There is no distortion of communication in order to fit known language.

COMPLEXITY

- Level 1 Texts may be simple showing little development. Simple sentences with little attempt at cohesion are acceptable.
- Level 2 Texts will display basic organization with themes and topics linked in a

simple way.

Level 3 Texts can be organized with themes and topics appropriately linked and sequenced. There will be a clear structure to the text where appropriate.

Level 4 There is clear and consistent evidence of the ability, to produce organized coherent and cohesive discourse where appropriate.

4. Other Types of Assessment

a. Portfolio

What are portfolios? We often ask ourselves or others this question because it is somewhat vague and unclear term in terms of authentic assessment. Now we are going to discuss more about it: the definition, forms and its format in assessing students' competence.

Definition

Applebee and Langer (1992, p.30) define portfolio as a cumulative collection of the work students have done. It can contain the student's total writing output, or a selection of works which the student has chosen for the teacher to evaluate.

Form

Some of the most popular forms of portfolio are the following:

1. a traditional "writing folder" in which students keep their work.
2. a bound notebook with separate sections kept for work in progress and final drafts.
3. a loose-leaf notebook in which students keep their draft and revisions.
4. a combination folder and big brown envelope where students' writings-exercises, tests, compositions, drafts, and so on-are kept.
5. a notebook divided into two sections: one for drafts and the other for final copies.
6. Reports on students' attitude towards certain subjects.

Portfolios provide a place for students to collect pieces that they believe mark their progress in the program. For example, if a student reads a book from cover to cover for the first time, he might take a picture of that book and put the picture in his portfolio. Students may put in their portfolios pieces of writing they have done, photocopies of articles they have read, or any other reading they might have done that held significance to them. They may include a letter they received in the portfolio. Any and all items or events that students believe are noteworthy and that reflect their progress may be put into or represented in the portfolio. (***Creating Authentic Materials and Activities for the Adult Literacy Classroom*** -Erik Jacobson, Sophie Degener Victoria Purcell-Gates).

An example of Recapitulation Format of Portfolio.

Recapitulation of Students' Assignment

Name : Class :
 NIS : Semester :

NO.	TYPES OF ASSIGNMENT	SK/ KD	SCORE	SIGNATURE		NOTE
				STUDENT	TEACHER	

Teacher's comment/note:

Jakarta,..... 2009

Classroom Teacher,

b. Self-Assessment

Learning Logs

Learning logs help teachers see what their students are learning, particularly in the writing class, and in the language class as a whole. In a learning log, students write on the knowledge they have gained from studying in their writing classes, and form their own thinking. A teacher need not grade learning logs, but can assess how much a student has gained or benefited from writing class.

Journal Entries

In here, the teacher can write short notes in response to students' thoughts. It can be a source of conferencing. It is enjoyable because students can write any topic at the spur of the moment, and it is also private.

Below is a comparison between learning log and journal entries according to Andrea H. Penaflores, (2005)

Dialogue Journal

- Teacher and student write to each other taking equal turns in writing and responding
- Teacher and student share ideas and Information
- Teacher and student act as equal partners in the interaction between them
- It is applicable to some content area courses such as literature, social studies, or science.
- teacher gives students assistance beyond what to do.

Journal Entries

- Teacher comments on students' work, but there is no equal turn taking in responding
- Student is not obligated her writing with anybody.
- There is a hierarchical relationship between teacher and student.
- Journal keeping is usually practiced in language courses only.
- Teacher assists students on the language used or on the content of what is written.

G. Test item Analysis

Teachers often assume that the test items which they have made are already good. As a result, if their students' scores are not satisfactory, they just take conclusion that it is their students who have not mastered the materials well while, in fact, the success of one test is determined not only by the students but also on the quality of the test item itself.

Therefore, keeping that in mind, teachers need to evaluate the quality of their tests items and analyze the test result as well as its construction. The procedure to find out the effectiveness of test item is called test item analysis (analisis butir soal). The already made test item is then tried out. The test data is analyzed in terms of its reliability and validity, as it is with the test item. The result of the analysis is revised (usually dropped out because of wording). Then, the result of revision is used for the real test.

Test item analysis is a test on the quality of the test item so that the characteristics of test item can be obtained.

Analysis of test item is done before use or tried out. The analysis includes: the

material, the construction, and the language

Analysis of empirical data of the result of testing is done afterward.

The analysis of test item can be carried out after the test is tried out to a number of students. Generally, there are two characteristics of item that can be viewed empirically, namely: the level of difficulty and the level of discrimination.

The Analysis of Multiple Choice Item.

The analysis of multiple choices would provide information about:

Index of facility of test item.

The level of discrimination of test item.

The effectiveness of each option.

From the analysis of the test item, the information about whether the item is too easy or too difficult can be acquired; whether the item can discriminate between students with high competence, and students with middle or low competence. It can also give information whether all the options can function effectively.

Steps to analyze multiple- choice items.

Here is the procedure to analyze multiple choice item:

- a. Put the result of the test in order; from the highest to the lowest score.
- b. Classify the test result into “Upper Group” for 27% students with highest test (starting from the highest score) and “Lower Group” for 27% students with the lowest score (starting from the lowest score).
- c. For each test item, count the number of students who choose each item for the upper and lower group.
- d. Do the step c above using the following format:

No. of test item	Group	Option					Number
		a	b*	c	d	o	
	A (Upper)						
	B (Lower)						
	Number						
	%						

Note: (*) is the correct answer

(o) is the number of students who do not choose any option

- e. Count the index of difficulty/easiness of each test item.
- d. Count the discrimination power of each item

Index of facility of Test Item (K)

Index of easiness of test item is the proportion of the test takers who answer correctly, and can be counted with the formula:

$$K = \frac{B}{T}$$

Note: K = index of easiness of test item

B = number of the test takers who answer correctly

T = number of the test takers

To count the index of facility of test item using the sample of upper and lower group, the formula is:

$$K = \frac{B_A + B_B}{n(A) + n(B)}$$

Note: B_A = The number of students from upper group who answer correctly

B_B = The number of students from lower group who answer correctly

$n(A)$ = the number of students in the upper group

$n(B)$ = the number of student in the lower group

The following is the criteria of interpreting one test item:

Facility index	Interpretation
$K < 0,30$	difficult
$0,30 \leq K \leq 0,70$	fair
$K > 0,70$	easy

Discriminating Power of Test items

Discriminating power is the ability of test item to discriminate between clever student (able to absorb the material) and less clever students (unable to absorb the material).

Index of discriminating power is the value that indicates the intensity of discriminating power of test item. To count the index of

discriminating power, count the different proportion between the students who answer correctly in the upper group (P_A) and the students who answer correctly in the lower group (P_B), the formula is :

$$D = P_A - P_B = \frac{B_A}{n(A)} - \frac{B_B}{n(B)}$$

If $n(A) = n(B)$, hence the following formula is:

$D = \frac{B_A - B_B}{n(A)} = \frac{B_A - B_B}{n(B)}$

Index of discriminating power is approximately between -1,00 and 1,00. The negative mark indicates that the item cannot discriminate between the clever and less clever students.

If the number of students who answer correctly in the upper group equals the number of students who answer correctly in the lower group, then the test item has $D = 0$. This might happen if the test is too easy so that every student can answer correctly, or the item is too difficult so that no one can answer correctly.

Criteria of Interpreting Discrimination Power

According to R.L. Ebel in his book entitled "Essentials of Educational Measurement", the criteria of interpreting discrimination power is as the following:

Discriminating power	Interpretation of test item
$D \geq 0,40$	good
$0,30 \leq D < 0,40$	fair
$0,20 \leq D < 0,30$	poor
$D < 0,20$	very poor

The effectiveness of Option

By doing the analysis of test item distribution of the test takers on one item, we can know about:

- a. the number of test takers who answer correctly.
- b. which distracter(s) that can function effectively.
- c. which distracter(s) that cannot function.
- d. which distracter(s) which are misleading.

A good distracter will attract more on the test-taker from the lower group than those in upper group.

The analysis of test taker distribution for one test item can use the following guidelines:

- a. Viewed from the number of the test taker who choose on the test sample, the distracter is said to function if:
 - 1) For five-option items, the distracter is chosen by at least 3% of all test takers in both upper and lower group.
 - 2) For four-option items, the distracter is chosen by at least 5% of all test takers in both upper and lower group.
- b. Viewed from the ones who choose distracters in upper and lower group.

If N_A = the number of test takers in upper group that choose distracters.

N_a = the number of test takers in lower group that choose distracters.

Then :

- 1) The distracter is said to be effective if $N_A < N_B$
- 2) The distracter is said to be misleading if $N_A > N_B$
- 3) The distracter is said to be ineffective if $N_A = N_B$

Below is an example of format of test analysis (multiple-choice) quoted from BINTEK KTSP.

Kartu telaah soal bentuk pilihan ganda

Aspek yang ditelaah	Nomor Soal							
	1	2	3	4	5	6	7	8
<p>Materi</p> <ol style="list-style-type: none"> 1. Soal sesuai dengan indikator (menuntut tes tertulis untuk bentuk pilihan ganda) 2. Materi yang diukur sesuai dengan kompetensi (urgensi, relevansi, kontinuitas, dan keterpakaian sehari-hari tinggi) 3. Pilihan jawaban homogen dan logis 4. Hanya ada satu kunci jawaban <p>Konstruksi</p> <ol style="list-style-type: none"> 1. Pokok soal dirumuskan dengan singkat, jelas, dan tegas 2. Rumusan pokok soal dan pilihan jawaban merupakan pernyataan yang diperlukan saja 3. Pokok soal tidak memberi petunjuk kunci jawaban 4. Pokok soal bebas dari pernyataan yang bersifat negatif ganda 5. Gambar, grafik, tabel, diagram, atau sejenisnya jelas dan berfungsi 6. Panjang pilihan jawaban relatif sama 7. Pilihan jawaban tidak menggunakan pernyataan "semua jawaban di atas salah/benar" dan sejenisnya 8. Pilihan jawaban yang berbentuk angka/waktu disusun berdasarkan urutan besar kecilnya angka atau kronologisnya 9. Butir soal tidak bergantung pada jawaban soal sebelumnya. <p>Bahasa/Budaya</p> <ol style="list-style-type: none"> 1. Menggunakan bahasa yang sesuai dengan kaidah bahasa Indonesia 2. Menggunakan bahasa yang komunikatif 3. Tidak menggunakan bahasa yang berlaku setempat/ tabu 4. Pilihan jawaban tidak mengulang kata/kelompok kata yang sama, kecuali merupakan satu kesatuan pengertian. 								

Analysis of Essay Item

Basically, the way to analyze essay item is the same as that of analyzing multiple choice item. The difference being is that there are only two ways of observing imperially, namely index of facility and the item's discriminating power.

Below is a sample of essay test item analysis quoted from BINTEK KTSP.

Analisis soal uraian dan tes praktek

NO.	SISWA	Soal 1 (Skor maks 6)	Soal 2 (Skor maks 5)
1	A	6	5
2	B	5	4
3	C	3	2
4	D	3	2
5	E	2	1
	Jumlah	19	14
	Rata-rata	3,80	2,80
	TK	0,63	0,56
	DP	0,47	0,56

TK1 = Rata-rata : skor maks

$$= 3,8 : 6 = 0,63$$

$$\text{TK2} = 2,8 : 5 = 0,56$$

DP1= (Rata-rata KA – Rata-rata KB) : skor maks.

$$= [(11:2) - (8:3)] : 6 = (5,5-2,7):6 = 0,47$$

$$\text{DP2} = [(9:2) - (5:3)] : 5 = (4,5-1,7) : 5 = 0,56$$

CHAPTER III

CLOSING REMARKS

This module of Language Assessment has covered some key terms related to assessment, some theoretical as well as practical aspects of assessment along with some samples from which teachers can develop their own ideas.

The assessment discussed in this supplementary module is classified into testing four language skills (Listening, Speaking, Reading, and Writing). Likewise, it is enriched with tasks after certain points are discussed to further deepen and develop teachers' understanding on the matter. Furthermore, test item analysis is also included in the hope that teachers can construct good test items.

Although this supplementary module is not presumably comprehensive enough to cover such wide subject as language assessment, it hopefully could provide assistance to teachers in understanding the main module in the BERMUTU training. Teachers can browse it to find certain item or points which they need for further comprehension or other specific needs like doing an assignment. Yet in the end, it is still up to the teacher to incorporate the relevant knowledge and skills gained from the training (and the modules) into their teaching practices in the framework of their professional development for upgrading quality education in Indonesia.

Suggestions and constructive criticisms will highly be appreciated.

REFERENCES:

www.yesdil.com

Bachman, L. F. 1990. Fundamental Consideration in Language Testing. Oxford: OUP

Brown, D.H. 2004. Language Assessment: Principles and Classroom Practices. London: Pearson Longman.

Depdiknas. (2002). Penilaian Berbasis Kelas. Jakarta: Puskur, Balitbang.

Heaton, J.B . 1988. Writing English Language Tests. ELBS.

Jacobson, E., Degener, S. and Victoria, P.G. Creating Authentic Materials and Activities for the Adult Literacy Classroom. Henning, G. 1987. A Guide to Language Teaching. Rowley: Newbury House Publishers.

Renandya, W. and Richards, J.C. (Editors) xxxx. Methodology In Language Teaching.

Panjaitan, M.O. xxxx Penilaian Pembelajaran Bahasa Inggris SMP.

Hughes, A. 2005. Testing for Language Teachers, Cambridge: CUP.

Renandya, W. 2005. Methodology In Language Teaching. Cambridge: CUP.

Weir, C.J. 1990. Communicative Language Testing. London: Prentice Hall.

SUGGESTED READING

Johnson, E.B. 200. Contextual Teaching & Learning. Diterjemahkan oleh Alwasilah, A.C.Bandung: MLC.

